Information Asymptotics and Inequalities for Posterior and Predictive Distributions

by

Robert McCulloch

University of Minnesota
School of Statistics
Technical Report No. 476
June 1986

Summary

We show that as the sample size goes to infinity the Kullback-Leibler divergence between the posterior and predictive distributions corresponding to different prior distributions goes to zero, almost surely. We also show that the Kullback-Leibler divergence between two posterior distributions is greater than the Kullback-Leibler divergence between the corresponding predictive distributions.

1. Introduction

It is well known that as the sample size gets large the importance of the prior distribution diminishes. Blackwell and Dubins(1962) show that almost surely as the sample size goes to infinity the L_1 norm between two predictive densities corresponding to two different priors goes to 0. Edwards, Lindman and Savage(1963) give conditions under which the L_1 norm between posteriors is small. These conditions will hold in large samples.

In this paper we show that the Kullback-Leibler divergence between posterior and predictive distributions which correspond to different priors goes to 0, almost surely, as the sample size goes to infinity . Kullback(1967) has shown that if the Kullback divergence between two densities goes to 0 then the L_1 norm goes to 0. Thus the result given in this paper is stronger than those just cited. However, the assumptions are stronger as well.

The relevance of the Kullback-Leibler divergence to statistical problem has been discussed by many authors. In particular Kullback(1959) shows that many statistical procedures may be motivated by the KLD. Also Akaike(1973) points out the very strong relation between the KLD and the likelihood. Bayesian statisticians use the Kullback-Leibler divergence in small sample procedures. See for example Johnson and Geisser(1983).

We also show that the KLD between two predictive densities is less than the KLD between the corresponding posterior distributions. This result is used in proving the asymptotic result but is also of interest in that it indicates that the predictive distribution is less sensitive than the posterior to specification of the prior.

2. Notation and Basic Assumptions

The asymptotics in this paper will be based on Fraser and McDunnough(1984), which will hereafter be referred to as FM. FM give a relatively simple and elegant proof of the asymptotic normality of posterior distributions based on reasonably simple and intuitively acceptable assumptions.

We now discuss notation and assumptions which will be identical to those of FM. For the convenience of the reader we reproduce the assumptions given in FM here.

Let $x_1, x_2, \ldots x_n$ be independent and identically distributed with density function $f(x|\theta)$, where the parameter θ takes on values in the real line R. The case where the parameter space is an open interval can be included by reparametrization. The observed likelihood function is given by

$$L_n(\theta) \propto f(x_1|\theta)\ldots f(x_n|\theta),$$

and the log-likelihood function by $l_n(\theta) = \ln L_n(\theta)$.

We shall denote the "true" value of θ by θ^*.

We use the following basic three assumptions throughout:

(I) $\limsup_{n \to \infty} \frac{1}{n} \sup_{\{\theta: |\theta - \theta^*| \geq \delta\}} [l_n(\theta) - l_n(\theta^*)] < 0$

for any $\delta > 0$, almost surely.

(II) $l_1(\theta)$ is twice continuously differentiable with

$0 < E(-\ddot{l}_1(\theta)) < \infty$.

Let $\sigma_n^{-2}(\theta) = E(-\ddot{l}_n(\theta))$. Note that $\sigma_n^{-2}(\theta) = n\,E(-\ddot{l}_1(\theta))$ so that $\sigma_n^{-2} \to \infty$.

(III) For each $\varepsilon > 0$, there exists $\delta > 0$ such that

$\limsup_{n \to \infty} \sigma_n^2(\theta^*) \sup_{\{\theta: |\theta - \theta^*| < \delta\}} |\ddot{l}_n(\theta) - \ddot{l}_n(\theta^*)| < \varepsilon.$

almost surely.

In the above the expectations are with respect to the distribution corresponding to θ^* and the almost sure refers to independent replications of this distribution.

For discussion of these assumptions see FM. The first ensures consistency of the maximum likelihood estimator, the second states that the Fisher information exists, and the third gives sufficient continuity of the second derivative of l_n with respect to θ.

In the following expressions $\int f(\theta)$ will mean the integral of f over the entire real line with respect to Lesbesgue measure.

Let $\hat{\theta}_n$ be the maximum likelihood estimator and $\hat{\sigma}_n^2 = \sigma_n^2(\hat{\theta}_n)$. Let θ_n be a real random variable having the probabilty density function $p_n(\theta) \equiv \frac{p(\theta)\,L_n(\theta)}{\int p\,L_n}$ where $p(\theta)$ is the prior density. θ_n has the posterior distribution.

If we let $T_n = \frac{\theta_n - \hat{\theta}_n}{\hat{\sigma}_n}$ then the probability density function corresponding to T_n is $h_n(t) = \dfrac{\hat{\sigma}_n \, p(\hat{\theta}_n + \hat{\sigma}_n t) \, L_n(\hat{\theta}_n + \hat{\sigma}_n)}{\int p \, L_n}$.

FM show that if p is continuous at θ^*, $p(\theta^*) > 0$, and $\int p \, L \leq \infty$ then $h_n(t) \to (1/\sqrt{2\pi}) \exp(-t^2/2)$ (the standard normal density) pointwise and uniformly for t belonging to a bounded set

almost surely. They also show that $\dfrac{\hat{\sigma}_n \, p(\hat{\theta}_n) \, L_n(\hat{\theta}_n)}{\int p \, L_n} \to (1/\sqrt{2\pi})$ almost surely, a result which we will find useful below. It follows from the above result that the distribution of T_n goes weakly to the standard normal distribution almost surely.

Thus the distribution of θ_n goes weakly to a point mass at θ^* since we have assumed that $\hat{\theta}_n$ is consistent. Consequently for any bounded continuous function g of θ, $\int g(\theta) \, p_n(\theta) \to g(\theta^*)$. However many functions of interest such as $g(\theta) = \theta$ are not bounded.

3. Asymptotic Results for Posteriors

Our first result modifies the proof in FM slightly to accomodate unbounded g. In all our work we assume I, II, and III above hold so that the results of FM are available.

<u>Result 1</u>: Suppose g is continuous at θ^* and the prior density p is continuous and postivive at θ^*.

Then, $\int p\, g\, L_1 \le \infty$ is sufficient for $\int g(\theta)\, p_n(\theta) \to g(\theta^*)$ almost surely as $n \to \infty$, and $\int g\, L_1 \le \infty$ is sufficient for

$$\int g(\hat{\theta}_n + \hat{\sigma}_n\, t)\ \ (1/\sqrt{2\pi})\ \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n\, t)}{L_n(\hat{\theta}_n)}$$

$$\to g(\theta^*) \text{ almost surely as } n \to \infty.$$

<u>Proof</u>:

Let $\theta = \hat{\theta}_n + \hat{\sigma}_n t$. Then

$\int g(\theta)\, p_n(\theta) = \int g(\hat{\theta}_n + \sigma_n t)\, h_n(t)$.

Let $A = [a,-a]$, $B_n = \{t: |\ \hat{\theta}_n + \hat{\sigma}_n t - \theta^*\ | < \delta\}$

Recall that assumptions I, II, and III imply that $\hat{\theta}_n \to \theta^*$ and $\hat{\sigma}_n \to 0$ almost surely.

$$\int g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t) = \int_A g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t)$$

$$+ \int_{A^c \cap B_n} g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t) \quad + \quad \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t).$$

We will show that for large a the first term is virtually $g(\theta^*)$ and that the other terms are negligable.

Almost surely there exists N such that for all $n \ge N$, $t \in A$, $|\ g(\hat{\theta}_n + \hat{\sigma}_n t) - g(\theta^*)\ | < \varepsilon$ by the continuity of g at θ^*.

So, for $n \ge N$, $|\ \int_A g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t) \ - \ g(\theta^*)\ |$

$$\le \int_A |\ g(\hat{\theta}_n + \hat{\sigma}_n t) - g(\theta^*)\ |\ h_n(t)$$

$$\leq \varepsilon \int_A h_n(t) \rightarrow \varepsilon\, (\, \Phi(a) - \Phi(-a)\,)$$

almost surely, as $n \rightarrow \infty$.

Here Φ stands for the standard normal c.d.f.

Now note that for all a , $(\, \Phi(a) - \Phi(-a)\,) = \leq 1$.

Hence, $\lim_{n \rightarrow \infty} \int_A g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t) = g(\theta^*)$ almost surely,

for all $a > 0$.

Now we need only show:

(i) $\lim_{a \rightarrow \infty} \lim_{n \rightarrow \infty} \int_{A^c \cap B_n} g(\hat{\theta}_n + \hat{\sigma}_n)\, h_n(t) = 0$ a.s.

(ii) $\lim_{n \rightarrow \infty} \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t) = 0$ a.s.

(i): For $t \in B_n$ we can make $g(\hat{\theta}_n + \hat{\sigma}_n t)$ and $p(\hat{\theta}_n + \hat{\sigma}_n t)$ as close as we like to $g(\theta^*)$ and $p(\theta^*)$ by our choice of δ. Thus it is sufficient to show that

$$\int_{A^c \cap B_n} \hat{\sigma}_n\, p(\hat{\theta}_n)\, \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{\int p\, L_n}$$

$$= \int_{A^c \cap B_n} c_n\, \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}$$

has the desired limiting property where

$$c_n \equiv \frac{\hat{\sigma}_n\, p(\hat{\theta}_n)\, L_n(\hat{\theta}_n)}{\int p\, L_n}, \quad c_n \rightarrow (1/\sqrt{2\pi}) \text{ a.s.}$$

But FM show that $\displaystyle \lim_{a\to\infty} \lim_{n\to\infty} \int_{A^c \cap B_n} \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}$

$= 0$ almost surely

so (i) holds.

(ii): Now FM show that $\dfrac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}$

$$\le \frac{L_1(\hat{\theta}_n + \hat{\sigma}_n t)}{L_1(\hat{\theta}_n)} \quad e^{-dn}$$

almost surely, for all $t \in B_n^c$, and some $d > 0$.

Hence, $\displaystyle \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t)\, h_n(t)$

$$= \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t) \frac{\hat{\sigma}_n \; p(\hat{\theta}_n + \hat{\sigma}_n t) \; L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{\int p\, L_n}$$

$$= \frac{c_n}{p(\hat{\theta}_n)} \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t)\, p(\hat{\theta}_n + \hat{\sigma}_n t) \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}$$

$$\le \frac{c_n \; e^{-dn}}{p(\hat{\theta}_n)} \int_{B_n^c} g(\hat{\theta}_n + \hat{\sigma}_n t)\, p(\hat{\theta}_n + \hat{\sigma}_n t) \frac{L_1(\hat{\theta}_n + \hat{\sigma}_n t)}{L_1(\hat{\theta}_n)}$$

$$\le \frac{c_n \; e^{-dn}}{p(\hat{\theta}_n)} \int_R g(\hat{\theta}_n + \hat{\sigma}_n t)\, p(\hat{\theta}_n + \hat{\sigma}_n t) \frac{L_1(\hat{\theta}_n + \hat{\sigma}_n t)}{L_1(\hat{\theta}_n)}$$

$$= \frac{c_n\ e^{-dn}}{p(\hat{\theta}_n)\ \hat{\sigma}_n\ L_1(\hat{\theta}_n)} \qquad \int g(\theta)\ p(\theta)\ L_1(\theta).$$

Now $(1/\hat{\sigma}_n)$ is $O(\sqrt{n})$ so $e^{-dn}/\ \hat{\sigma}_n \to 0$, the integral in less than infinity by assumption and $c_n/(p(\hat{\theta}_n)\ L_1(\hat{\theta}_n))$ is bounded for n large as long as L_1 is bounded away from 0 for θ near θ^*. Thus (ii) holds and the result is proved. The second part of the result follows from a similar argument. □

We use the above result to show that the Kullback divergence between two posteriors corresponding to two different priors goes to 0 as the sample size goes to infinity.

We start with a lemma.

<u>Lemma 1</u>: Suppose p and q are two prior densities which are continuous and positive at θ^* and both $\int p\ L_1$ and $\int q\ L_1$ are less than infinity. Then $(\int p\ L_n)/(\int q\ L_n) \to p(\theta^*)/q(\theta^*)$ almost surely.

<u>Proof</u>: As above we assume that the parameter space is the real line. Let $\theta = \hat{\theta}_n + \hat{\sigma}_n t$. Then,

$$\frac{\int p\, L_n}{\int q\, L_n} = \frac{\int p(\hat{\theta}_n + \hat{\sigma}_n t)\, L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{\int q(\hat{\theta}_n + \hat{\sigma}_n t)\, L_n(\hat{\theta}_n + \hat{\sigma}_n t)}$$

$$= \frac{\int p(\hat{\theta}_n + \hat{\sigma}_n t)\, \frac{1}{\sqrt{2\pi}}\, \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}}{\int q(\hat{\theta}_n + \hat{\sigma}_n t)\, \frac{1}{\sqrt{2\pi}}\, \frac{L_n(\hat{\theta}_n + \hat{\sigma}_n t)}{L_n(\hat{\theta}_n)}}$$

$\rightarrow\ p(\theta^*)/q(\theta^*)$ almost surely by the previous result. □

Result 2: Suppose p and q are two prior densities which are continuous and positive at θ^*. Suppose $\int p\, L_1$, $\int q\, L_1$, and $\int \frac{p^2}{q} L_1$ are all $\leq \infty$. Then the Kullback divergence between p_n and q_n goes to 0 almost surely, where $p_n = p\, L_n \,/ \int p\, L_n$, $q_n = q\, L_n \,/ \int q\, L_n$, and the Kullback divergence between p_n and q_n equals $\int p_n \log(p_n \,/\, q_n)$.

Proof:

It is easy to show that $\log(x) \leq x - 1$ so,

$$\int p_n \log(p_n / q_n) \leq \int p_n\, (p_n / q_n - 1)$$

$$= \int p_n^2 / q_n - 1.$$

So it sufficient to show $\int p_n^2 / q_n \rightarrow 1$.

Now, $p_n^2 / q_n = (p_n / q_n)\, p_n$

$$= \frac{\int q\, L_n}{\int p\, L_n}\ p/q\ p_n\ .$$

So, $\int p_n^2 / q_n = \dfrac{\int q\, L_n}{\int p\, L_n} \int p/q\ p_n$.

Let $g(\theta) = p(\theta) \,/\, q(\theta)$.

$\int g\, p\, L_1 = \int p^2/q\, L_1 \le \infty$ by assumption. We have also assumed that p and q are continuous and non zero at θ^* so that g is continuous and non zero at θ^*. Thus by result 2 above, we have $\int p/q\, p_n = \int g\, p_n$

$$\rightarrow g(\theta^*) = p(\theta^*)/q(\theta^*).$$

By the lemma 2 above $\dfrac{\int q\, L_n}{\int p\, L_n} \rightarrow q(\theta^*)/p(\theta^*)$.

$$\text{Hence, } \int p_n^2 \,/\, q_n = \frac{\int q\, L_n}{\int p\, L_n} \int g\, p_n \rightarrow \frac{q(\theta^*)}{p(\theta^*)}\ \frac{p(\theta^*)}{q(\theta^*)} = 1 ,$$

and the result is proved. □

4. An Inequality

In this section we prove an inequality which relates predictive distributions to corresponding posterior distributions. The basic intuition behind the result is that if you take two different posteriors, arising from two different priors, the corresponding predictive distributions tend to be "closer together" than the posteriors. We use the Kullback-Leibler divergence to measure how close together distributions are. The mathematical tools used are Jensen's inequality and Fubini's theorem.

The result of this section tells us that, given a choice for the distribution of the observable quantity conditional on the parameter value, the predictive distribution is less sensitive to the choice of the prior than the posterior distribution. Thus if the predictive distribution is the desired output less care is needed in choosing the prior than if the posterior distribution is the output. On the other hand, there is a sense in which the predictivist has a tougher job than someone making inferences about parameters. Since he makes statements about variables which are potentially observable his claims may be directly refuted by experience.

Many authors have stressed the importance of the predictive distribution. While parameters may in some cases be useful devices for describing the structure of a model, their values and the model itself are ultimately only of interest if they make useful predictions about potentially observable quantities. See for example Geisser(1971), Geisser(1980), and Zellner(1985).

We develop the inequality of this section in a more general framework than that of the previous sections. We now introduce the notation that will be used throughout this section.

First we have the parameter space Θ with sigma field B. All probability measures on Θ will be absolutely continuous with respect to the measure ν and hence representable by densities with respect to ν. As usual this underlying structure is summarized by the triplet (Θ,B,ν).

Second we have the observation space Z. Again we have an underlying structure for probability statements summarized by (Z,C,μ) where C is a sigma field and μ is a measure on (Z,C). The Z space lists the possible observable outcomes as apposed to the Θ space which lies entirely within the realm of our imagination (or beyond).

Third we have the model which specifies a distribution on Z for each $\theta \in \Theta$. The model is described by means of a function $f:\Theta\times Z \to R$ with $f(z|\theta)$ being a density with respect to μ for each θ. Since many of the arguments use Fubini's theorem we note the product space $(Z\times\Theta, C\times B, \mu\times\nu)$. We assume that f is measurable with respect to the space $(Z\times\Theta, C\times B)$.

Let p be a density for a probability distribution on Θ. We then let $f_p(z) = \int_\Theta f(z|\theta)\, p(\theta)\, d\nu(\theta)$. If we are thinking of p as a posterior distribution on Θ we often call f_p the density of the predictive distribution. If p is the density for a prior on Θ then f_p is the no-data predictive density or simply the marginal density where we think of the model f and the prior p as having induced a joint distribution on $(Z\times\Theta, C\times B)$. Note that f_p is C measureable.

<u>Result 3</u>: Let $a(z) = \int_\Theta f(z|\theta)\, d\nu(\theta)$. If $0 < a(z) < \infty$ for all $z \in Z$ then $K(f_p, f_q) \le K(p, q)$ where K is the Kullback divergence.

<u>Proof</u>:

If $K(p,q) = \infty$ then the result is obvious.

Hence we may assume that $K(p,q) < \infty$.

Now $p\log(p/q) = p(-\log(q/p))$

$$\ge p(1 - (q/p))$$

$$= p - q .$$

Also $\int_\Theta (p - q)^- d\nu \le \int_\Theta | p - q | d\nu \le \int_\Theta (p + q) d\nu = 2 < \infty$

so $(p\log(p/q))^- \le (p - q)^-$ and $\int_\Theta (p\log(p/q))^- d\nu < \infty$.

Thus $K(p,q) = \int_\Theta p\log(p/q)\, d\nu < \infty$ implies $\int_\Theta | p\log(p/q) | d\nu < \infty$.

We now use $\int_\Theta | p\log(p/q) | d\nu < \infty$ to show that

$$\int\int_{Z\times\Theta} | f(z|\theta)\, p(\theta)\log(p(\theta)/q(\theta)) | \, d\nu\times\mu(\theta,z) < \infty$$

a fact which justifies a future use of fubini's theorem.

$$\int\int_{Z\times\Theta} | f(z|\theta)\, p(\theta)\log(p(\theta)/q(\theta)) | \, d\nu\times\mu(\theta,z)$$

$$= \int\int_{Z\times\Theta} f(z|\theta)\, | p(\theta)\log(p(\theta)/q(\theta)) | \, d\nu\times\mu(\theta,z)$$

$$= \int_\Theta \int_Z | p(\theta)\log(p(\theta)/q(\theta)) | \, f(z|\theta)\, d\mu(z)\, d\nu(\theta)$$

(applying Fubini to a positive function)

$$= \int_\Theta | p\log(p/q) | \, d\nu < \infty.$$

We now proceed to the main part of the proof.

$$f_p(z) = \int_\Theta f(z|\theta)\, p(\theta)\, d\nu(\theta) = a(z) \int \frac{f(z|\theta)}{a(z)}\, p(\theta)\, d\nu(\theta)$$

$$\equiv a(z)\, E_z(p(\theta))$$

Where E_z means expectation with respect to the probability measure on (Θ,B) having density $\frac{f(z|\theta)}{a(z)}$ supported by the measure ν. Similarly we have $f_q(z) = a(z)\, E_z(q(\theta))$.

Now for ψ convex on R_+^2 we have $\psi(E_z p, E_z q) \le E_z \psi(p,q)$ by an application of Jensen's inequality.

The function $\psi(x,y) = x\log(x/y)$ is convex on R_+^2 so $(E_z p)\log((E_z p)/(E_z q)) \le E_z\, p\log(p/q)$, which implies

$$\frac{f_p(z)}{a(z)} \log\left(\frac{f_p(z)}{f_q(z)}\right) \leq \int_\Theta \frac{f(z|\theta)}{a(z)} \; p(\theta)\log(p(\theta)/q(\theta)) \; d\nu(\theta)$$

$$f_p(z)\log\left(\frac{f_p(z)}{f_q(z)}\right) \leq \int_\Theta f(z|\theta) \; p(\theta)\log((p(\theta)/q(\theta)) \; d\nu(\theta)$$

So $\int_Z f_p(z) \; \log(f_p(z)/f_q(z)) \; d\mu(z) \leq$

$$\int_Z \int_\Theta f(z|\theta) \; p(\theta)\log(p(\theta)/q(\theta)) \; d\nu(\theta) \; d\mu(z)$$

$$= \int_\Theta \int_Z p(\theta)\log(p(\theta)/q(\theta)) \; f(z|\theta) \; d\mu(z) \; d\nu(\theta)$$

$$= \int_\Theta p\log(p/q) \; d\nu \qquad \square$$

5. Asymptotics for Predictive Distributions

Our final result is an immediate consequence of the last two. The Kullback divergence between the posteriors goes to 0 almost surely and the Kullback divergence between the predictives is less than that between the posteriors. Hence the Kullback divergence between the predictives goes to 0. Note that the f of section 4 which relates θ to distributions on the observable quantity need not be the same of the f of the previous sections which relates the sample information to the parameter θ providing us with the posterior. The quantities x_i which make up our sample need not even be restricted to the same set as the quantities which are trying to predict. Since we refer to result from all the previous sections let $F(z|\theta)$ be the denstity of the distribution for the quantity z

which we wish to predict and $f(x_i|\theta)$ be the density of our sample quantities.

Result 4: Assume that the conditions of Result 2 hold and that $\int F(z|\theta)\, d\theta$ is between 0 and ∞ for all z in the prediction space. Let $F_n^p(z) = \int F(z|\theta)\, p_n(\theta)$ and $F_n^q(z) = \int F(z|\theta)\, q_n(\theta)$, where, as before p_n and q_n are the posteriors corresponding to the priors p and q given a sample of size n.
Then the Kullback divergence between F_n^p and F_n^q goes to 0 almost surely as $n \to \infty$.

proof:

Let $K(f,g) = \int f \log(f/g)$, the Kullback divergence.

By result 3 we have, $K(F_n^p,F_n^q) \le K(p_n,q_n)$. By result 2 $K(p_n,q_n) \to 0$ almost surely as $n \to \infty$. Hence,

$$0 \le K(F_n^p,F_n^q) \le K(p_n,q_n) \to 0 \text{ almost surely.} \qquad \square$$

6. Conclusion

The framework of Fraser and McDunnough provides a powerful tool for Bayesian asymptotics. Note that while the development of this paper assumes θ to be a real number the results easily extend to the case where θ is a vector. The results may also be easily extended to the case where the observations are not independent. Again, both of these extentions are straightforward because Fraser and McDunnough provide simple extensions of their results to these cases.

Acknowledgment

The author thanks Professor Seymour Geisser for many helpful comments. This work was funded by the Social Sciences and Humanities Research Council of Canada and NIH grant NIGMS 25271.

References

Akaike, H. (1973). Information theory and an extention of the maximum likelihood principle. 2-nd International Symposium on Information Theory. Edited by Petrov and Csaki. Publishing House of the Hungarian Academy of Sciences. 267-281.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. Annals of Math. Stat. 33, 82-886.

Edwards, W., Lindman, H., and Savage, L. (1963). Bayesian statistical inference for psychological research. Psychological Review. 70, 193-242.

Fraser, D.A.S. and McDunnough, P. (1984). Further remarks on asymptotic normality of likelihood and conditional analyses. Canadian J. of Stat. 12, 183-190.

Geisser, S. (1971). The inferential use of predictive distributions Foundations of Statistical Inference. Edited by Godambe and Sprott. Toronto, Holt, Rinehart, and Winston. 456-466.

Geisser, S. (1980). A predictivistic primer. Bayesian Analysis in Econometrics and Statistics. chapter 24, edited by Arnold Zellner. North Holland.

Johnson, W. and Geisser, S. (1983). A predictive view of the detection and characterization of influential observations in regression analysis. J. Amer. Statist. Assoc. 78, 137-144

Kullback, S. (1959). Information Theory and Statistics. New York Wiley.

Kullback, S. (1967). A lower bound for discrimination information in terms of variation. I.E.E.E. Transactions on Information Theory. 126-127.

Zellner, A. (1985). Bayesian analysis in econometrics. Tech. Report, H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago.